SF 298 MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	09/25/98	Final Report

4. TITLE AND SUBTITLE

Elastic Spider Silk Proteins

5. FUNDING NUMBERS

DAAH04-95-1-0531

6. AUTHOR(S)

Randolph V. Lewis

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(ES)

University of Wyoming
Laramie, WY 82071

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

ARO 34417.4-LS

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Flagelliform silk proteins have been studied by cloning the cDNAs for the major protein in that silk as well as the gene. The protein consists of three sequence segments which compose a repeat. These repeats appear numerous times in the protein. The three segments are: 1) GPGGX; 2) GGX; and 3) a highly conserved non-silk-like "spacer" sequence. The gene is composed of the same repeats combined with a highly conserved intron.

Two major publications have been published and four patents awarded during the tenure of this grant.

14. SUBJECT TERMS

Spider silk, sequencing, flagelliform, Nephila clavipes

15. NUMBER IF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Final Report for DaaH04-95-1-0531

4.

A. Spider silks have become widely recognized for their mechanical properties, which are comparable or surpass other natural fibers and even many manmade materials. For example, some silks exhibit reversible stretch greater than high elastic nylon (>200%) while others have a tensile strength (>200,000 psi) greater than steel. The molecular basis for how spider silks achieve such feats was the focus of our research. In particular, we 1) cloned and sequenced cDNA for the previously uncharacterized flagelliform silk; 2) determined the genomic organization of the flagelliform silk gene; 3) developed hypotheses relating silk sequence motifs to specific mechanical properties; and 4) designed a strategy to study the expression and fiber formation of flagelliform silk protein.

B. (1) Characterization of Flagelliform Silk cDNA

We accomplished the primary goal of cloning the gene for flagelliform silk (Hayashi & Lewis, 1998), the stretchiest of all known silks. Flagelliform silk forms part of capture spiral of an orb-web and has a lower tensile strength ($1x10^9$ Nm^{-2}) but several times the extensibility (>200%) of dragline silk (Vollrath & Edmonds, 1989; Köhler & Vollrath, 1995). A functioning capture spiral is a composite of secretions from the aggregate and flagelliform silk glands. We focused on flagelliform silk because it forms the actual fiber of the spiral while aggregate silk is laid down as a non-fibrous, aqueous coating of sticky droplets.

The mechanical properties of flagelliform silk correspond to the capture spiral's ecological function. A spider's orb-web has to immediately stop a rapidly flying insect in a manner that allows the prey to become entangled and trapped. To do this the web must absorb the energy of the insect without breaking and yet not act as a trampoline to bounce the insect away from the web. Without the high elasticity of the capture spiral, an orb-web could not be as effective an aerial net.

To clone the flagelliform silk protein gene, we constructed a cDNA library from mRNA expressed in the flagelliform silk gland of *Nephila clavipes*. From this library we generated complete sequences for several cDNA clones. These clones span ~6000 basepairs of the flagelliform gene and include a putative secretory signal, the carboxy-terminal end, and substantial portions of intervening sequence (see Genbank accession numbers AF027972 and AF027973). When this large amount of sequence data is translated, it becomes clear that the flagelliform protein can be summarized by three repeating amino acid repeats: Gly-Pro-Gly-Gly-X; Gly-Gly-X; and a non-glycine rich "spacer." This protein surprisingly lacks the poly-Ala regions that are prevalent in *Bombyx* (the silkworm) silk and the other characterized spider silks.

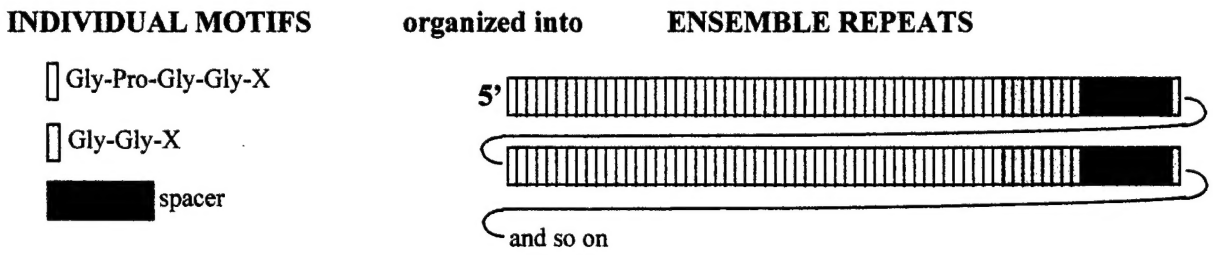
The Gly-Pro-Gly-Gly-X motif is the dominant repeat. The first four amino acids of this unit are highly conserved among the numerous individual repeats. The fifth amino acid, indicated by X, is variable. However, only a small subset of residues (Ala, Ser, Tyr, and Val) occupies 90% of those positions. Moreover, the distribution of the four residues among the repeats is non-random. Repeat units with X=Ala tend to be followed by other units with X=Ala. This pattern creates an array of exact Gly-Pro-Gly-Gly-Ala repeats. A second type of array has repeats with X=Tyr alternating with repeats with X=Ser or Val. Flagelliform is the only silk protein known to have these types of patterned higher level variation.

The second glycine-rich motif, Gly-Gly-X, occurs approximately tenfold fewer times than Gly-Pro-Gly-Gly-X. Similar to the proline containing motif, the X residue is predominantly Ala or Ser. Gly-Gly-X is present between the Gly-Pro-Gly-Gly-X arrays and the spacer and may serve as a transition between them.

The third repetitive element is both the longest and least common motif. These regions are termed "spacers" because they disrupt the glycine and proline rich flagelliform sequence. Though at 28 amino acids in length, each spacer is substantially longer than a Gly-Pro-Gly-Gly-X or Gly-Gly-X motif, the spacers are extremely highly conserved. The individual spacers differ by only one or two residues from each other.

Not only is the flagelliform sequence composed of individual repetitive motifs, but the motifs themselves are organized into larger ensemble repeats. The Gly-Pro-Gly-Gly-X motifs are arrayed in tandem forty-three to sixty-three times. These repeats are followed by six to twelve Gly-Gly-X motifs, a "spacer" region, and finally a single Gly-Gly-X. Then, the ensemble repeat begins again with Gly-Pro-Gly-Gly-X.

Repetitive Units Within the Flagelliform Silk Protein



We propose that the three repetitive motifs and their occurrence in ensemble repeats reveals the structural basis for elasticity. The dominant motif of this protein, Gly-Pro-Gly-Gly-X, appears up to 63 times in tandem arrays. This motif likely forms Pro^2-Gly^3 type II β-turns with the resulting series of concatenated β-turns forming a helix (termed a β-spiral:). This β-spiral, much like a spring, can be stretched and upon release of tension, recoils back to its original length.

This simple model provides for elasticity at the level of individual β-spirals. However, the existence of distinct Gly-Pro-Gly-Gly-X neighborhoods (i.e. X=Ala arrays or X=Tyr alternating with X=Ser/Val arrays) suggests that interactions between β-spirals also have significance. Research on elastins (Urry et al. 1995), glutens (Van Dijk et al., 1997), and the bacterial virulence factor P.69 pertactin (Emsley et al., 1996) has shown that bonds can form between adjacent β-spirals. As the flagelliform β-spirals associate, the protein monomers can align and assemble into silk fibers.

The spacer regions may also be involved in the assembly of silk fibers. While the precise structure and function of the spacers remains unknown, their remarkably high sequence conservation and possession of the only negatively

charged residues (Asp, Glu) among the repetitive units suggests that it is a critical component of the silk. We propose that the spacers create distinct regions where the flagelliform proteins can overlap and align with each other, resulting in a fiber of woven monomers. The negatively charged residues of the spacers may also promote post-assembly interactions with the coating of aqueous aggregate gland silk.

(2). Genomic Organization of the Flagelliform Silk Gene

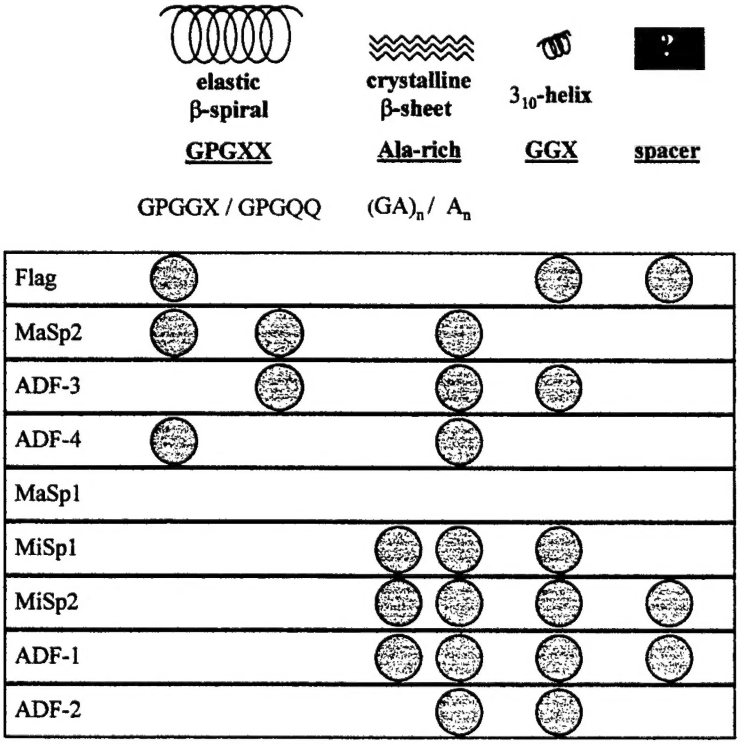
In addition to cloning cDNAs for the flagelliform protein, we also isolated a large portion of the silk gene from *Nephila clavipes* genomic DNA through a combination of screening a λ library and PCR amplification (Hayashi & Lewis, in prep.). We completely sequenced the ~17.6 kb of DNA fragments and confirmed the cDNA results. More importantly, the genomic sequences reveal that flagelliform has a unique gene organization. The gene is evenly divided between coding and non-coding sequence. Aside from a small exon of non-repetitive 5' sequence, all the exons encode a single ensemble repeat (the higher level repeat unit made up of Gly-Pro-Gly-Gly-X, Gly-Gly-X, and spacer subunits). Even more surprising was the discovery that the introns between these iterative exons also share an extremely high level of identity. Thus, the introns and exons themselves are an even higher level of repeating unit within the flagelliform gene. This gene organization provides strong evidence that concerted evolution is involved in the maintenance and diversification of flagelliform silk.

(3). Hypotheses Relating Silk Amino Acid Motifs to Mechanical Properties

One of the attractive features of studying spider silks is that unlike other silk-producing organisms, spiders produce multiple types of silks during all stages of their lifetime. Because each silk is used for a specific ecological function, the different silks have their own distinctive combinations of mechanical properties. This naturally occurring diversity allows for comparative studies among closely related proteins. Thus we compared the new flagelliform data to the available

sequence and structural studies done for other spider silks (Hayashi, Shipley, & Lewis, submitted). We concentrated primarily on correlating amino acid motifs with two mechanical features: elasticity and tensile strength. The hypotheses that were generated have significance for the design of synthetic silks with properties precisely engineered for specific applications.

The three best known silks differ in their elasticity and strength. Major ampullate (dragline) silk has moderate elasticity and high tensile strength while minor ampullate (web reinforcement) silk has a lower tensile strength and lacks elasticity. Flagelliform silk also has a lower tensile strength than major ampullate silk but has much higher elasticity. The sequencing of cDNAs for these silks has established that they are composed almost entirely of repetitive elements. By comparing the consensus repeats from the different silks, we have shown that there are four types of amino acid motifs shared by all known spider silks: 1) poly-Ala/poly-Gly-Ala; 2) GGX; 3) GPGGX/GPGQQ; and 4) "spacers." All of these elements were also found orthologous cDNAs from *Araneus* (Guerette et al., 1996).



abbreviations: Flag=flagelliform protein; MaSp1, MaSp2, ADF-3, ADF-4=major ampullate proteins; MiSp1, MiSp2, ADF-1=minor ampullate proteins; ADF-2=putative tubuliform protein

Evidence from biophysical studies on silks suggest that specific secondary structures are related to some of these shared motifs. Fiber X-ray diffraction and NMR data have been most useful with substantiating the presence of poly-Ala and poly-Gly-Ala regions as β-sheet (Simmons et al., 1994; Kümmerlen et al., 1996; Simmons et al., 1996; Parkhe et al. 1997). These regions could serve as the linkage points for the crystalline areas in the fiber. Presumably, these are the parts of the protein that bind the monomers together in the fiber and provide tensile strength. These β-sheets can be depicted with poly-Ala forming a structure with successive alanine residues placed on alternate sides of a backbone. Each chain can then interlock with an adjacent chain. This configuration provides additional hydrophobic binding energy as the β-sheet regions are known to be poorly hydrated (Matsuno & Lewis, unpub. data).

The poly-Gly-Ala regions can form a structure similar to poly-Ala. The key feature of the poly-Gly-Ala configuration is that all the glycines are on one side of the backbone and all the alanines are on the other side. Because the glycine side of the polypeptide chain is unable to have the same number of hydrophobic interactions possible with poly-Ala, the poly-Gly-Ala regions have a lower binding energy than poly-Ala β-sheets. This model is in agreement with the lower tensile strength of minor ampullate silk (with poly-Gly-Ala) relative to major ampullate silk (with poly-Ala). Thus, the strength of interactions between the β-sheet regions of these proteins predicts the tensile strength of each silk.

The second shared motif is Gly-Gly-X. These repeat regions have been proposed to form either a β-sheet (Thiel et al., 1994) or a helix. We prefer the helical motif because it is supported by both FTIR and NMR data (Dong et al., 1991; Kümmerlen et al., 1996). A tight 3_{10} helix is consistent with GGX being three residues in length. Such helices could serve as a transition or link between crystalline β-sheet regions and less rigid protein structures. Also, neighboring GGX helices may interact to maintain alignment among adjacent protein molecules in the fiber.

The GPGXX pentapeptide repeat, the third shared motif, has been suggested to conform to a structure similar to the β-turn spiral of elastin (Urry et al., 1975; Chang et al., 1989; Urry et al., 1995). Our particular model for the β-spiral formed by the GPGGSGPGGY segment of flagelliform silk is shown at right. Two features are notable in this model. The first is the similarity of this structure to a spring that could easily serve as the elastic mechanism in the fiber. The proline residue would be the focal point for the retraction energy after stretching.

By forcing the proline bonds to torque in response to extension, a large force can be generated for retraction. The second key feature of the model is the positioning of the hydroxyls in Ser and Tyr for hydrogen bonding with downstream Gly residues. Notice that the long Tyr sidechains stabilize the tight β-turns and the shorter Ser sidechains stabilize the layers of coils. The importance of these bonds is supported by the strong tendency for Tyr and Ser residues to regularly alternate in flagelliform silk.

Only the major ampullate and flagelliform silks contain the GPGXX motif and they are also the stretchiest of spider silks. As further support of the GPGXX motif providing the elasticity module, there is a correspondence between the number of tandemly arrayed GPGXX repeats and the different extensibilities of the two silks. *Nephila* major ampullate silk, with up to 35% extension has at most nine β-turns in a row before interruption by another motif (Hinmam & Lewis, 1992). Flagelliform silk with 200% extensibility has a minimum of 43 contiguously linked β-turns in its spring-like spirals (Hayashi & Lewis, 1998). Thus, the longer the molecular spring, the greater the elasticity of the silk.

The fourth shared motif, the "spacers," are the most complex repeats in silks. The structures formed by these regions remain unknown. Though the minor ampullate (Colgin & Lewis, 1998) and flagelliform (Hayashi & Lewis, 1998) spacers differ radically in amino acid sequence, they are both relatively long and contain charged residues. Possible roles for the spacers include: 1) pre-fiber--promotion of an alternative structure for the silk while it is stored in liquid form to prevent

premature fiber formation within the silk glands; 2) within fiber--alignment of crystalline or other structural regions among individual protein molecules; and 3) extra-fiber--provision of surface regions that interact with other critical components of silk (e.g., the association of flagelliform silk with sticky aggregate gland secretions that are important for trapping prey).

Based on the associations presented above relating amino acid motifs to secondary structures and secondary structures to functions in the silk fiber, we view spider silks as sets of modules. These modules, such as the crystalline module poly-Ala, provide specific properties to a silk fiber. Not only is the presence of a particular module important, but the frequency of modules is critical. Thus, the greater extensibility of flagelliform silk compared to major ampullate silk can be attributed to flagelliform having more repeats of the elasticity module GPGXX. Similarly, the greater tensile strength of major ampullate silk compared to minor ampullate silk is due to the large poly-Ala component in major ampullate silk and the presence of the weaker poly-Gly-Ala crystalline regions in minor ampullate silk.

These modular hypotheses both explain the properties of the known silks and predict the properties of silks yet to be characterized. It may be that spider silks have evolved through the modification and shuffling of a small number of amino acid motifs.

(4). Recombinant Protein and Expression Studies

Based on our previous success with artificial gene construction using nonregenerable restriction sites (Lewis et al., 1996), we designed DNA cassettes to code for $(GPGGYGPGGS)_2$ and $(GPGGA)_4$. These can be combined to generate sequences which are similar to the native protein. We can also construct GGX repeats and the spacer sequences as well. Initial efforts are directed toward the first two cassettes to aid in determining the structure of the proteins in the fiber.

References

Chang, D., C. Venkatachalam, and D. Urry. 1989. Nuclear Overhauser effect and computational characterization of the beta-spiral of the polypentapeptide of elastin. *J. Biomol. Struct. Dyn.*, 6:851-858.

Colgin, M. and R. Lewis. 1998. Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like "spacer regions." *Protein Science*, 7:667-672.

Dong, Z., R. Lewis, and C. Middaugh. 1991. Molecular mechanisms of spider silk elasticity. *Arch.Biochem.Biophys.*, 284:53-57.

Emsley, P., Charles, I., Fairweather, N. and Isaacs, N. 1996. Structure of *Bordetella pertussis* virulence factor P.69 pertactin. *Nature, 381*, 90-92.

Guerette, P., D. Ginzinger, B. Weber, and J. Gosline. 1996. Silk properties determined by gland-specific expression of a spider fibroin gene family. *Science*, 272:112-115.

Hayashi, C., N. Shipley, and R. Lewis. submitted to *International Journal of Biological Macromolecules*. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silks.

Hayashi, C. and R. Lewis. 1998. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J.Mol.Biol.*, 275:773-784.

Hinman, M. and R. Lewis. 1992. Isolation of a clone encoding a second dragline silk fibroin. *J.Biol.Chem.*, 267:19320-19324.

Köhler, T. and F. Vollrath. 1995. Thread biomechanics in the two orb-weaving spiders *Araneus diadematus* (Araneae, Araneidae) and *Uloborus walckenaerius* (Araneae, Uloboridae). *J. Exp. Zool.* 271:1-17.

Kümmerlen, J., J. van Beek, F. Vollrath, and B. Meier. 1996. Local structure in spider dragline silk investigated by two-dimensional spin-diffusion NMR. *Macromolecules*, 29:2920-2928.

Lewis, R., M. Hinman, S. Kothakota and M. Fournier. 1996. Expression and Purification of a Recombinant Spider Silk Protein: A New Strategy for Producing Repetitive Proteins. *Protein Expression and Purification* 7:400-406.

Parkhe, A., S. Seeley, K. Gardener, L. Thompson, and R. Lewis. 1997. Structural studies of spider silk proteins in the fiber. *J.Mol.Recognition*, 10:1-6.

Simmons, A., E. Ray, and L. Jelinski. 1994. Solid state ^{13}C NMR of *N.clavipes* dragline silk establishes structure and identity of crystalline regions. *Macromolecules*, 27:5235-5237.

Simmons, A., C. Michal, and L. Jelinski. 1996. Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science*, 271:84-87.

Thiel, B., D. Kunkel, and C. Viney. 1994. Physical and chemical microstructure of spider dragline: a study by analytical transmission electron microscopy. *Biopolymers*, 34:1089-1097.

Urry, D., T. Onishi, M. Long, and L. Mitchell. 1975. Studies on the conformation and interactions of elastin: nuclear magnetic resonance of the polyhexapeptide. *Int. J. Pept. Protein Res.*, 7:367-378.

Urry, D., C. Luan, and S. Peng. 1995. Molecular biophysics of elastin structure, function and pathology. *Ciba Found.Symp.*, 192:4-30.

Van Dijk, A., De Boef, E., Bekkers, A., Van Wijk, L., Van Swieten, E., Hamer, R. and Robillard, G. 1997. Structure characterization of the central repetitive domain of high molecular weight gluten proteins. II. Characterization in solution and in the dry state. *Protein Science*, 6, 649-656.

Vollrath, F. and D. Edmonds. 1989. Modulation of the mechanical properties of spider silk by coating with water. *Nature*, 340:305-307.

C. Publications

(1995) Colgin, M. and Lewis, R.V. Spider Silk: An Ancient Biomaterial for the Future. Chemistry and Industry 24:1009-1012.

(1996) Lewis, R.V., Hinman, M., Kothakota, S. and Fournier, M. Expression and Purification of a Recombinant Spider Silk Protein: A New Strategy for Producing Repetitive Proteins. Protein Expression and Purification 7:400-406.

(1997) Parkhe, A.J., Seeley, S.K., Gardner, K., Thompson L. and Lewis, R.V. Structural Studies of Spider Silk Proteins in the Fiber. J. of Molecular Recognition 10:1-6.

(1997) Liivak, O., Flores, A., Lewis, R.V. and Jelinski, L.W. Conformation of the Polyalanine Repeats in Minor Ampullate Gland Silk of the Spider *Nephila clavipes*. Macromolecules 30:7127-31.

(1998) Colgin, M and Lewis, R.V. Spider Minor Ampullate Silk Proteins Contain New Repetitive Sequences and Highly Conserved Non-silk-like "Spacer Regions." Protein Science 7:667-672.

(1998) Hayashi, C. and Lewis, R.V. Spider Flagelliform Silk Proteins. J. of Mol. Biol. 275:773-784.

D. Participating personnel

Mark Colgin, graduate student (PhD. 1996)

Tony Flores, graduate student (M.S. expected 1999)

Tony Contento, graduate student

Nikki Shippley, graduate student

Cheryl Hayashi, postdoctoral

Mike Hinman, postdoctoral

7 undergraduate students

5. Inventions

A new cloning method for large DNA (#5,695,971) with Yoichi Kadokami

Spider Silk Protein (#5,728,810) with Mike Hinman and Ming Xu

cDNAs Encoding Minor Ampullate Spider Silk Proteins (#5,733,771) with Mark Colgin

Minor Ampullate Spider Silk Proteins (#5,756,677)

Spider silk protein sequences and structure (2 more currently pending)